2. Analysis of Product Reviews (multiligual) + User Interface (Unilever)

Mentor Information

Mentor Name: Dr. Ye Xu
Title: Sr. Research Scientist
Company: Unilever

Motivation and Background

As consumer purchases continue to shift from brick-and-mortar stores to e-commerce, free-text product reviews is becoming an increasingly important reservoir of consumer insight. For consumer goods companies like Unilever, such reviews – whether generated via online retailers or social media sites or through product testing prior to launches – can hold the key to developing the next revolutionary product. Numerous market research and analytics companies are already playing in this space, offering solutions with varying levels of automation. Being a global company, Unilever's consumers reside all over the world and speak many different languages, and we have not come across a solution that can be utilized across our key markets for our questions of interest. We did a Capstone project with students in Sept. 2016. The students successfully finished the first part, which was the summary and key word counting in English. However students do not have time to finish it for different language and the user interface. This is the continue effort of the last Capstone project.

The Data Set

A data set containing approx. 200,000 reviews of beauty products from Amazon.com for training & algorithm development purposes, and several internal datasets from Unilever consumer tests across the world in various languages.

Project Overview

How much can we automate the process of analyzing free responses from consumers around the world? This project covers two key objectives: 1. Development of a text analysis methodology or tool that can extract and quantify the popularity of key themes and associations from free response datasets. 2. Assess whether automated translation APIs (e.g. Microsoft Translator) can produce outputs of sufficient quality for this type of insight extraction from free text data. If not, explore types of preprocessing, analyses, and/or algorithms can mitigate issues caused by poorer quality translations or integrated approach from data scraping from social media site to analysis in multiple languages to interpretation etc.

Research Goals

1. Using a product review dataset for training and development, provide free-text, sentiment analysis functions for extracting key themes and associations between concepts, as well as quantifying their popularity and strength. Generate consumer insights from the product review dataset. We recommend that the analysis account for typos, synonyms, and abbreviations where possible.
2. Provide integration of an automated translation service/API with the free-text analysis functions and demonstrate effectiveness on Unilever consumer test datasets, or provide clear explanation on why current auto translations are of insufficient quality.
3. Recommend next steps for development.

Suggested Output

1. Slide deck summarizing the methodology and key insights from the analyses in this project.
2. Algorithms/code or a compiled executable for performing the analysis with well-documented instructions.
3. Recommended methodology or function for extracting summary insight from the analysis.

Product Follow-Ons (e.g., apps)

An integrated software or Excel add-on with the text analysis functionalities